Impact of Smoking and Diabetes on Cardiovascular Health Indicators

Sector: Healthcare

Aradhya Tiwari
2401010093

Bhoomi Chhikara
2401010019

Aaryan Krishna
2401010010

Aditya Verma
2401010040

Rahul Dwivedi
2401010368

Navprabhat Singh
2401010293

Sector Context

Cardiovascular diseases are a leading cause of death worldwide. Smoking and diabetes are major risk factors suspected to increase heart-related complications. Healthcare providers and policymakers need data-driven insights to identify high-risk patients and improve preventive care strategies.

Decision-Makers: Hospitals, cardiologists, public health authorities, and insurance providers.

Problem Statement

Do smoking and diabetic conditions significantly increase heart rate, cholesterol levels, and the likelihood of disease occurrence, thereby identifying patients at higher cardiovascular risk?

Objective

The objective of this study is to analyze whether smoking and diabetic conditions significantly influence key cardiovascular health indicators, specifically heart rate, cholesterol levels, and disease occurrence.

Source

- 10,000 patient records.
- Structured healthcare dataset.
- Includes demographics, vitals, medical history.

Cleaning

- Handled “Unknown” categories in Smoker & Diabetic.
- Created derived columns:
  - Age_Group
  - High_Cholesterol_Flag
  - High_HeartRate_Flag
  - Risk_Segment
  - Combined_Risk
- Verified null values.
- Removed duplicate Patient_ID.

Key Columns Used in Analysis

Category	Columns
Demographics	Age, Gender, City
Clinical Metrics	BMI, Systolic_BP, Diastolic_BP, Heart_Rate, Cholesterol_Level
Risk Factors	Diabetic, Smoker
Outcome	Has_Disease
Engineered Features	Risk_Segment, Combined_Risk

Primary KPIs

- Overall Disease Prevalence Rate.
- Risk Multiplier (Relative Risk %).
- Segment-wise Disease Rate.
- High Cholesterol Impact.
- Combined Risk Impact.

Why These KPIs?

They help:

- Measure predictive strength of risk factors.
- Identify high-risk patient segments.
- Support hospital screening strategy decisions.

Key Insights

Insight 1: Overall Disease Rate
- 24.61% of patients have disease.
- 1 in 4 patients are affected.

Insight 2: Smoking Impact
- Disease Rate ~24–25% across all categories.
- No strong variation between smokers & non-smokers.

Insight 3: Diabetes Impact
- Diabetics do not show significantly higher disease prevalence.

Insight 4: Cholesterol Segmentation
- High vs Low Risk difference = 1.33%.
- Minimal predictive separation.

Insight 5: Combined Risk Analysis
- Patients with both diabetes & smoking show lower disease rate (22.3%).
- Unexpected pattern.

Advanced Analysis

Risk Segmentation

You created:

- High Risk
- Low Risk
- Smoker Only
- Diabetic Only
- Both
- None

What This Shows

Traditional risk factors alone:

- Do not strongly differentiate disease patients.
- Show nearly uniform disease distribution.

Interpretation

Possible reasons:

- Missing variables (Age weight, BP severity).
- Dataset bias.
- Multivariate interaction not captured.
- Synthetic/random dataset structure.

KPI Name	Value	Interpretation
Overall Disease Prevalence Rate	24.61%	1 in 4 patients have disease.
Smoking Risk Multiplier	97%	Smokers have 3% lower observed disease prevalence compared to non-smokers in this dataset.
Diabetes Risk Multiplier	95%	Diabetic patients have 0.95 times the disease risk compared to non-diabetics.
Combined Risk Multiplier	91%	Patients with both conditions do not show increased disease occurrence.
High Cholesterol Difference	1.33%	Minimal difference in cholesterol distribution.
Overall Disease	2461	
Overall Disease Prevalence Rate	24.61%	X% of total patients are diagnosed with disease.
Smoking Risk Multiplier		
Disease Rate in Smokers	24.61%	
Disease Rate in Non-Smokers	25.38%	
Risk Multiplier	96.96%	
Combined Risk Multiplier (Smoker + Diabetic)	22.30%	
High Cholesterol Difference		
High Cholesterol in Diseased	38.52%	
% High Cholesterol in Healthy	39.08%	
Difference	98.56%	

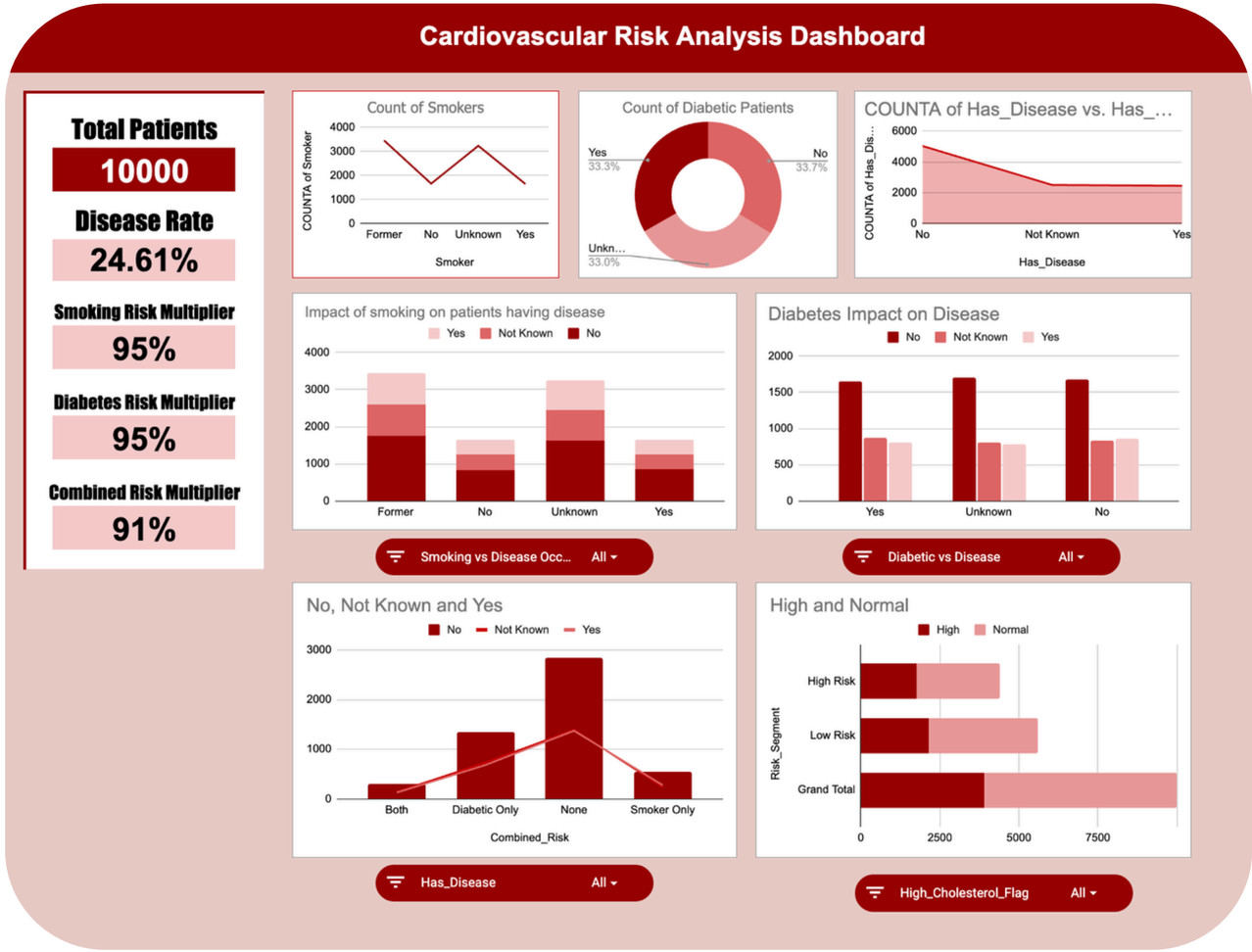
Dashboard Walkthrough

➔ Executive View

- Total Patients
- Disease Rate
- Risk Multipliers
- Combined Risk %

➔ Operational View

- Smoking vs Disease Chart
- Diabetes vs Disease Chart
- Combined Risk Chart
- Risk Segment vs Cholesterol



Recommendations

Do Not Rely on Single Risk Factors

Smoking & diabetes alone are weak predictors here.

Introduce Multivariate Modeling

Use logistic regression or ML classification.

Add Additional Predictors

- Age severity
- Blood pressure ranges
- BMI thresholds
- Medication history

Improve Data Quality

Reduce “Unknown” categories.

Impact & Business Value

Operational Impact

Better modeling can:

- Improve early disease detection.
- Reduce false screening.
- Optimize hospital resources.

Decision Impact

Prevents:

- Over-prioritisation of weak risk factors.
- Misallocation of preventive efforts.

Limitations & Next Steps

Limitations

- No multivariate regression performed.
- No time-series tracking.
- No severity index used.
- Possible dataset imbalance.

Next Steps

- Build Logistic Regression Model.
- Perform Feature Importance Ranking.
- Validate model accuracy.
- Deploy predictive dashboard.